

# A NEW MONTE CARLO METHOD FOR ASSESSING TAXONOMIC SIMILARITY WITHIN FAUNAL SAMPLES: REANALYSIS OF THE GASTROPOD COMMUNITY OF ONEIDA LAKE, NEW YORK

ROBERT T. DILLON, JR.  
DEPARTMENT OF BIOLOGY, COLLEGE OF CHARLESTON,  
CHARLESTON, SOUTH CAROLINA 29424, U.S.A.

## ABSTRACT

Using frequency table analysis and log-linear models, Dillon (1981) concluded that bottom samples taken by F. C. Baker (1918) from Oneida Lake, New York, had significantly fewer pairs of confamilial snail species than expectation based on a Monte Carlo simulation unweighted by relative abundance. If confamilial species are assumed to have similar ecological requirements, these findings suggest that competition has played a role in determining the micro-distribution of snails in Oneida Lake. However, the statistical tests employed in 1981 were weak in many respects. So in this study, I propose a new method of assessing the taxonomic similarity within faunal samples to re-examine F. C. Baker's data. Samples are categorized simultaneously by the number of species and the number of higher taxa they contain using a tabular format, and the resulting distribution of samples by species is used in a Monte Carlo simulation. Results were similar to those of 1981. The taxonomic similarity of snail samples cannot be distinguished from random expectation based on an abundance-weighted model. But if species are assumed to have equal chances of occurring in samples, regardless of their relative abundances, samples from Oneida Lake tend to have substantially fewer genera and families than expected.

The similarity of co-occurring animals has been the object of considerable study and debate for about 40 years. The extensive literature has recently been reviewed by Harvey *et al.* (1983) and by Strong *et al.* (1983). In general, it has been established that a relationship exists between an organism's diet and its morphology. The more similar a pair of organisms are morphologically, the more likely it is that they will rely on similar resources. Thus early workers (Elton, 1946; Hutchinson, 1959) expected that co-occurring animals ought to be unusually dissimilar morphologically in order to reduce competition. Others (e.g. Simberloff, 1970) have suggested the opposite, that co-occurring animals may tend to be unusually similar, since similar animals have similar dispersal capabilities and similar ecological needs. Much debate has centered upon the statistical tests that can be appropriate to distinguish these two alternatives from a third, that no pattern exists at all regarding species similarities and distributions.

Two general methods have been used to estimate overall morphological similarity. The more direct approach involves measuring the size and shape of various anatomical features on representative specimens from each taxon being studied (Strong *et al.*, 1979; Simberloff and Boecklen, 1981; Bowers and Brown, 1982; Case *et al.*, 1983; Travis and

Ricklefs, 1983; Schum, 1984). Difficulties arise, however, in the selection of relevant characters to measure and appropriate individuals to measure them on. This latter problem is particularly acute in species (e.g. most mollusks) where there is no discrete adult size. Thus there are attractions to the use of taxonomic "relatedness" as a measure of morphological similarity (Elton, 1946; Williams, 1947; Simberloff, 1970). Here it is assumed that species in the same genus, for example, are very similar to each other. But species in different genera of the same family are somewhat less similar, the species of different families are less similar still, and so on. Data of this sort are very easy to obtain, but are somewhat difficult to analyse.

Dillon (1981) used both morphometric and taxonomic methods to estimate the similarity of snails co-occurring in small samples taken from the bottom of Oneida Lake, New York, by Baker (1918). Taxonomic similarity was estimated using the number of congeneric and confamilial pairs of species. Then the observed taxonomic similarities were compared to those expected from Monte Carlo simulations using frequency table analysis. But this method was weak in several respects. Because it was based on chi-square statistics, a great deal of data-pooling was necessary to obtain the minimum sample sizes required in each cell. Congeneric

triplets and quadruplets were difficult to handle. And further, the contribution of any particular factor to the fit eventually obtained between actual data and log-linear model cannot be assessed independently of other effects in frequency table. A number of indirect tests suggested, however, that some differences between the taxonomic similarity observed in Baker's data and that expected from simulations were substantial.

Here I describe a new test to analyse taxonomic similarity within faunal samples that avoids the difficulties outlined above. Instead of counting congeneric or confamilial pairs, entire distributions of genera or families are compared. I will use this new technique to reanalyse Baker's data on the distribution of gastropods in Oneida Lake.

## METHODS

Details regarding the collection of the data to be analysed here can be obtained in Baker (1918). Briefly, Baker made 162 quantitative samples of plants and macrobenthos, primarily using a long-handled dipper or a dredge. Twenty-one of these samples either contained no snails or were omitted from the report. Collected in the remaining 141 samples were 5,716 individual snails, representing 37 species and subspecies. Omitting very rare species and lumping those that have been synonymized, Dillon (1981) reduced these numbers to 5,582 individuals representing 23 species. The species involved, their distributions and abundances, and the higher systematic categories recognized are all given in Dillon (1981).

The 121 samples with more than one species present were first categorized simultaneously by the number of species and genera they contained. This was most conveniently accomplished using a data table with the number of species listed down the left margin and the number of genera listed across the top. Then the number of samples containing two species, three species, and so forth, was totalled down the right-hand margin of the table. The total number of samples containing one genus, two genera, and so forth, was totalled at the bottom. Distributions of samples by the number of species and higher taxa they contained will be referred to as S distributions and  $T_0$  distributions (higher taxa observed), respectively. Table 1 illustrates this technique. An identical procedure was also used to tabulate the samples by the number of families they contained.

If there is no tendency for co-occurring snails to be more or less similar to one another taxonomically, a random sample of species from the Oneida Lake fauna using the S distribution should give a distribution of genera or families ( $T_e$ , higher taxa expected) indistinguishable from  $T_0$ . But if co-occurring snails tend to be taxonomically dissimilar, for example, the  $T_0$  distribution will tend to be higher than the randomly-generated  $T_e$  distribution. Just as in the 1981 analysis,  $T_e$  distributions were obtained using two algorithms.

For the abundance-weighted test, a pool was created in which each snail species was represented according to its abundance over all 141 samples taken. For example,

**Table 1.** Baker's (1918) samples from Oneida Lake, New York, categorized by the number of species and genera of snails they contained. The row totals constitute the S distribution, and the column totals the  $T_0$  distribution.

	NUMBER OF GENERA								T
	1	2	3	4	5	6	7	8	
2		27							27
3			23						23
4			2	19					21
5			1	4	22				27
Number of	6			1	1	7			9
Species	7			1	1	3	4		9
	8					1			1
	9					1	1	1	3
	10						1		1
T	0	27	26	25	24	13	5	1	121

Baker collected a total of 17 *Campeloma decisum* (Say) in his 141 samples, so the probability of selecting *C. decisum* from the species pool was  $17/5,582 = 0.003$ . Notice that data from samples containing only one species are included in the calculation of relative abundances, although not in the compilation of the S distribution. Then a uniform random number generator was used to draw "samples" from the species pool, with replacement, following the S distribution. The number of random samples taken was 100 times the number of actual observations. For example, Table 1 shows that the S distribution has 27 samples with two species represented, 23 samples with three species, and so on, up to one sample with ten species. Thus in the computer simulation, 2700 samples were taken including two different species from the species pool, 2300 samples were taken of three different species, and so on, up to 100 samples of ten species. These randomly-generated samples, categorized by the number of genera of families they contained, constituted the two  $T_e$  distributions. Table 2 illustrates this method and shows the results from the analysis of genera.

Techniques were quite similar for the abundance-unweighted simulation, the only difference being that all 23 species had equal probabilities of being selected from the pool. Thus the probability of drawing *Campeloma decisum* was  $1/23 = 0.043$ . The two  $T_e$  distributions, one for genera and the other for families, were generated by drawing 100 times the S distribution as before. Copies of the computer program (in Basic) used for the generation of both weighted and unweighted  $T_e$  distributions are available from the author.

The  $T_0$  and  $T_e$  distributions were compared using values of the Kolmogorov-Smirnov statistic D from one-sample tests (Siegel 1956: 47). The D statistic is the maximum difference between the cumulative expected distribution and the cumulative distribution actually observed. Normally, D statistics are presented as absolute values. But for this application, a positive value of D will indicate that  $T_0$  distributions tend to take higher values than  $T_e$ , and therefore that co-occurring snails tend to be taxonomically dissimilar. A negative value of D will suggest the opposite. It should be cautioned that D-statistics are sensitive to any sort of deviation from expectation, not just difference in central tendency.

**Table 2.** Results of the Monte Carlo simulation of Baker's (1918) samples from Oneida Lake. The row totals are the S distribution, and the column totals the  $T_e$  distribution of genera.

		NUMBER OF GENERA									
		1	2	3	4	5	6	7	8	9	T
Number of Species	2	44	2656								2700
	3		154	2146							2300
	4		1	328	1771						2100
	5			40	797	1863					2700
	6				43	377	480				900
	7				3	124	476	297			900
	8					2	26	51	21		100
	9						32	120	122	26	300
	10						2	23	54	21	100
		T	44	2811	2514	2614	2366	1016	491	197	47

Thus the data were always plotted and examined critically before any conclusions were drawn from the D-statistics.

Ideally, one would want to know the likelihood that a  $T_o$  distribution might arise as a random sample from a given  $T_e$  distribution. The unusual composition of T distributions, however, precludes inference regarding the significance of D or any other conventional statistic. Although T distributions can theoretically take any frequency from 0.0 to 1.0 at the lower end of the scale, frequencies are constrained at values above 2 higher taxa present. Because no more than two higher taxa can be present when only two species are present, and no more than three higher taxa can be present in samples of three species, and so forth, T distributions are not completely free to vary at the upper end of their ranges. Thus it seems possible that  $T_o$  distributions would be more likely to underestimate than overestimate  $T_e$  distributions. That is, this technique would seem to be biased towards finding that co-occurring animals seem to be more similar than random expectation.

In order to investigate the strength of this and other potential biases, Dillon and Schotland (unpublished data) used this technique to analyse a large series of randomly-generated data sets. We found substantial bias only under very extreme conditions. In the normal range of species abundances and aggregations, there is little detectable difference between  $T_o$  and  $T_e$ . So although I can present no confidence estimates with the results of my analysis, simple inspection of D statistics and graphed results should give a reasonably reliable indication of trends in taxonomic similarity.

## RESULTS

The four comparisons between observed and expected taxonomic similarity are plotted in Figure 1. The observed data seem to fit abundance-weighted expectation fairly well. Values of D are 0.017 for the genus comparison and -0.083 for the family comparison. As a yardstick, the critical value of D from a one-sample K-S test with  $N = 121$  is 0.123 (two-tailed). Thus the probability that gastropod species co-occur in Oneida Lake would seem to be a function of relative abundances but not taxonomy. There is no evidence that con-

generic or confamilial species have significant tendencies to occur together or to occur apart, assuming the abundance-weighted hypothesis.

On the other hand, both  $T_o$  distributions seem to be shifted substantially to the right of  $T_e$  distributions based on abundance-unweighted simulations. The values of D are 0.107 for the genus comparison and 0.099 for the family comparison. Given the sample size of 121, these values are as large or larger than the most extreme values of D generated in the simulation tests of Dillon and Schotland. Thus there is fairly good evidence that snails co-occurring in samples taken from the bottom of Oneida Lake tend to be more dissimilar taxonomically than random expectation unweighted by species abundance.

## DISCUSSION

Although derived using a different technique, these results agree well with those of Dillon (1981). The earlier analysis also suggested that the taxonomic similarity of co-occurring snails seems to be indistinguishable from random expectation if the probability of occurrence for each species is weighted by its abundance. But if all species are equally likely to occur, it appears from both analyses that co-occurring snails tend to be taxonomically dissimilar.

Unweighted Monte Carlo simulations would initially seem to be less realistic and thus less interesting to test than the abundance-weighted ones. But if relative abundances are viewed as a function of recent environmental conditions and the life cycles of the species involved, these abundances can change rapidly. Thus abundance-unweighted "null hypotheses" have been more commonly tested by previous researchers.

Dillon (1981) examined the morphometric similarity of co-occurring gastropods as well as their taxonomic similarity. Judging from size and shape of the shell and radula, it was concluded that snail species co-occurring in Oneida Lake tend to be significantly more dissimilar than the abundance-weighted simulation would suggest. Considered along with the results of this investigation, these findings constitute some of the strongest published evidence of dissimilarity in co-

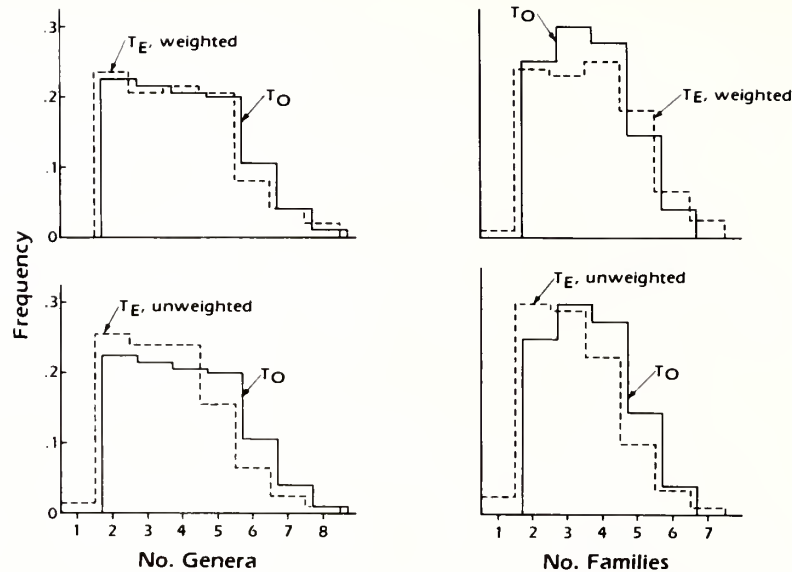


Fig. 1. Comparison of observed ( $T_O$ ) and expected ( $T_E$ ) distributions of gastropod samples from Oneida Lake, New York, by the number of higher taxa they contained. The  $T_E$  distributions are distinguished by a dashed line and are offset slightly from the  $T_O$  distributions.

occurring animals. Most workers (Simberloff, 1970; Strong *et al.*, 1979; Ricklefs and Travis, 1980; Ricklefs *et al.*, 1981; Simberloff and Boecklen, 1981) have found greater than expected similarity in samples of co-occurring animals.

But competition is only one of several possible explanations for the Oneida Lake results. For example, suppose that a pair of congeneric species are found to occupy different habitats, say sandy bottom and rocky bottom, such that they rarely co-occur. It could be that one species competitively excludes the other, or that the two species have adapted to different habitats as a response to competition in the past. Or it could be that the two species have diverged from a single ancestral species that previously occupied both bottom types, and competition has never played a role. Statistical tests such as the one described here are but a preliminary step towards the understanding of a very complex question.

### ACKNOWLEDGMENTS

I thank Tom Schotland for his expert computer programming. Computer time was furnished by the Department of Malacology, Academy of Natural Sciences of Philadelphia. The preparation of the manuscript and figures was supported by the Department of Biology, College of Charleston.

### LITERATURE CITED

- Baker, F. C. 1918. The productivity of invertebrate fish food on the bottom of Oneida Lake, with special reference to mollusks. *New York State College of Forestry Technical Publication No. 9*. 296 pp.
- Bowers, M. A. and J. H. Brown. 1982. Body size and coexistence in desert rodents: chance of community structure? *Ecology* 63:391-400.
- Case, T. J., J. Faaborg and R. Sidell. 1983. The role of body size in the assembly of West Indian bird communities. *Evolution* 37:1062-1074.
- Dillon, R. T., Jr. 1981. Patterns in the morphology and distribution of gastropods in Oneida Lake, New York, detected using computer-generated null hypotheses. *American Naturalist* 118:83-101.
- Elton, C. 1946. Competition and the structure of ecological communities. *Journal of Animal Ecology* 4:127-136.
- Harvey, P. H., R. K. Colwell, J. W. Silvertown, and R. M. May. 1983. Null models in ecology. *Annual Review of Ecology and Systematics* 14:189-211.
- Hutchinson, G. E. 1959. Homage to Santa Rosalia, or Why are there so many kinds of animals? *American Naturalist* 93:145-159.
- Ricklefs, R. E., D. Cochran, and E. R. Pianka. 1981. A morphological analysis of the structure of communities of lizards in desert habitats. *Ecology* 62:1474-1483.
- Ricklefs, R. E., and J. Travis. 1980. A morphological approach to the study of avian community organization. *Auk* 97:321-338.
- Schum, M. 1984. Phenetic structure and species richness in North and Central American bat faunas. *Ecology* 65:1315-1324.
- Siegel, S. 1956. *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, New York. 312 pp.
- Simberloff, D. S. 1970. Taxonomic diversity of island biotas. *Evolution* 24:23-47.
- Simberloff, D. S. and W. Boecklen. 1981. Santa Rosalia reconsidered: size ratios and competition. *Evolution* 35:1206-1228.
- Strong, D. R., D. S. Simberloff, L. G. Abele, and A. B. Thistle, eds. 1983. *Ecological Communities: Conceptual Issues and the Evidence*. Princeton University Press, Princeton, New Jersey. 613 pp.
- Strong, D. R., L. A. Szyska, and D. S. Simberloff. 1979. Tests of community-wide character displacement against null hypotheses. *Evolution* 33:897-913.
- Travis, J. and R. E. Ricklefs. 1983. A morphological comparison of island and mainland assemblages of neotropical birds. *Oikos* 41:434-441.
- Williams, C. B. 1947. The generic relations of species in small ecological communities. *Journal of Animal Ecology* 17:11-18.